52. IWK

Internationales Wissenschaftliches Kolloquium
International Scientific Colloquium

PROCCEDINGS | **10 - 13 September 2007**

FACULTY OF COMPUTER SCIENCE AND AUTOMATION

COMPUTER SCIENCE MEETS AUTOMATION

VOLUME II

TECHNISCHE UNIVERSITÄT
ILMENAU

Bibliografische Information der Deutschen Bibliothek
Die Deutsche Bibliothek verzeichnet diese Publikation in der deutschen Nationalbiografie; detaillierte bibliografische Daten sind im Internet über **http://dnb.ddb.de** abrufbar.

ISBN 978-3-939473-17-6

Impressum

Herausgeber:	Der Rektor der Technischen Universität Ilmenau Univ.-Prof. Dr. rer. nat. habil. Peter Scharff
Redaktion:	Referat Marketing und Studentische Angelegenheiten Kongressorganisation Andrea Schneider Tel.: +49 3677 69-2520 Fax: +49 3677 69-1743 e-mail: kongressorganisation@tu-ilmenau.de
Redaktionsschluss:	Juli 2007
Verlag:	 Technische Universität Ilmenau/Universitätsbibliothek Universitätsverlag Ilmenau Postfach 10 05 65 98684 Ilmenau www.tu-ilmenau.de/universitaetsverlag
Herstellung und Auslieferung:	Verlagshaus Monsenstein und Vannerdat OHG Am Hawerkamp 31 48155 Münster www.mv-verlag.de
Layout Cover:	www.cey-x.de
Bezugsmöglichkeiten:	Universitätsbibliothek der TU Ilmenau Tel.: +49 3677 69-4615 Fax: +49 3677 69-4602

© Technische Universität Ilmenau (Thür.) 2007

Diese Publikationen und alle in ihr enthaltenen Beiträge und Abbildungen sind urheberrechtlich geschützt. Mit Ausnahme der gesetzlich zugelassenen Fälle ist eine Verwertung ohne Einwilligung der Redaktion strafbar.

Preface

Dear Participants,

Confronted with the ever-increasing complexity of technical processes and the growing demands on their efficiency, security and flexibility, the scientific world needs to establish new methods of engineering design and new methods of systems operation. The factors likely to affect the design of the smart systems of the future will doubtless include the following:

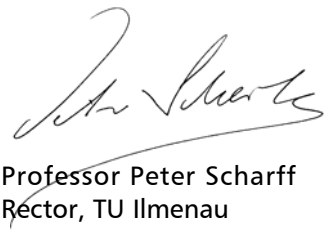
- As computational costs decrease, it will be possible to apply more complex algorithms, even in real time. These algorithms will take into account system nonlinearities or provide online optimisation of the system's performance.
- New fields of application will be addressed. Interest is now being expressed, beyond that in "classical" technical systems and processes, in environmental systems or medical and bioengineering applications.
- The boundaries between software and hardware design are being eroded. New design methods will include co-design of software and hardware and even of sensor and actuator components.
- Automation will not only replace human operators but will assist, support and supervise humans so that their work is safe and even more effective.
- Networked systems or swarms will be crucial, requiring improvement of the communication within them and study of how their behaviour can be made globally consistent.
- The issues of security and safety, not only during the operation of systems but also in the course of their design, will continue to increase in importance.

The title "Computer Science meets Automation", borne by the 52nd International Scientific Colloquium (IWK) at the Technische Universität Ilmenau, Germany, expresses the desire of scientists and engineers to rise to these challenges, cooperating closely on innovative methods in the two disciplines of computer science and automation.

The IWK has a long tradition going back as far as 1953. In the years before 1989, a major function of the colloquium was to bring together scientists from both sides of the Iron Curtain. Naturally, bonds were also deepened between the countries from the East. Today, the objective of the colloquium is still to bring researchers together. They come from the eastern and western member states of the European Union, and, indeed, from all over the world. All who wish to share their ideas on the points where "Computer Science meets Automation" are addressed by this colloquium at the Technische Universität Ilmenau.

All the University's Faculties have joined forces to ensure that nothing is left out. Control engineering, information science, cybernetics, communication technology and systems engineering – for all of these and their applications (ranging from biological systems to heavy engineering), the issues are being covered.

Together with all the organizers I should like to thank you for your contributions to the conference, ensuring, as they do, a most interesting colloquium programme of an interdisciplinary nature.
I am looking forward to an inspiring colloquium. It promises to be a fine platform for you to present your research, to address new concepts and to meet colleagues in Ilmenau.



Professor Peter Scharff
Rector, TU Ilmenau

Professor Christoph Ament
Head of Organisation

Table of Contents

C O N T E N T S

10 Mobile Communications

11 Education in Computer Science and Automation

M. Pfaff, D. Woetzel, D. Driesch, S. Toepfer, R. Huber, D. Pohlers, D. Koczan,
H.-J. Thiesen, R. Guthke, R.W. Kinne

Gene Expression Based Classification of Rheumatoid Arthritis and Osteoarthritis Patients using Fuzzy Cluster and Rule Based Methods

7 NEW METHODS AND TECHNOLOGIES FOR MEDICINE AND BIOLOGY

With gene expression data becoming increasingly available in medicine there is an increasing need to adequately analyse this kind of data and use it for biomarker identification and classification with respect to specific diseases in order to support medical diagnosis.

We report here on an investigation into the use of gene expression information for the classification of rheumatoid arthritis (RA) and osteoarthritis (OA) patients applying fuzzy cluster and rule based methods.

The aim of this investigation was to identify gene expression patterns that can be used to classify RA and OA patients.

RA and OA both belong to the chronic rheumatic diseases. Whereas RA represents an inflammatory joint disease with an aggressive, joint destructive character (affecting approximately 1% of the population in industrialised countries), OA is a degenerative rheumatic disease with superimposed inflammatory flares.

Current medical diagnosis of the two diseases is primarily based on clinical data. In the future, this diagnosis may be supported by a gene expression based diagnosis. Such an approach may not only aid future differential diagnosis, but also provide clues to disease mechanisms and an improved therapy.

The study described here focused on the application of fuzzy cluster and rule based methods to reveal interrelations between gene expression in synovial tissue and the clinical diagnosis of RA and OA as well as a control group (CG; neither RA nor OA, mostly joint trauma or normal).

The original database used in part in this study was assembled at the Rudolf Elle Hospital Eisenberg, Germany, and covers 268 patients (107 RA, 118 OA, 36 CG and 7 other). Gene expression data was available for 33 of these patients (13 RA, 10 OA, 10 CG) and for 22,283 Affymetrix® U133A gene fragments each.

The data analysis approach applied consists of five consecutive steps: data pre-processing, clustering, rule extraction, rulebase construction with optimisation as well as cluster and rule based classification with validation.

Data pre-processing included logarithm calculation and median based normalisation of the gene expression intensities as well as outlier detection and removal.

Clustering of the data into two clusters (low and high expression of the respective gene) was performed applying a modified fuzzy c-means algorithm.

Extraction of uni-conditional rules from the clustered data was carried out using a modified *Kiendl* relevance index to rate and rank the relevant rules with the conclusions RA, OA, CG.

Rulebases were then constructed and optimised using the highest ranked rules and used for the classification of patients with respect to the diseases/control.

The obtained results were finally subjected to leave-one-out cross-validation.

The rule extraction step yielded three ranked rule lists consisting of 45, 27, 7 rules with the conclusion CG, RA, OA, respectively.

Figure 1 shows the main results of the analysis. Patients are represented in the rows and genes in the columns. Genes, i.e. uni-conditional rules with the respective gene in the conditional part, are arranged from left to right for the respective patient group (CG, RA, OA) according to the *Kiendl* relevance index as calculated in the rule extraction step (highest *Kiendl* values left). The figure only includes relevant genes, i.e. those with a *Kiendl* relevance index >0. The patients are arranged from top to bottom according to the clinical diagnosis (CG, RA, OA; see also Status). For each patient and each relevant gene, the fuzzy membership value calculated in the preceding fuzzy c-means clustering step is plotted (‘Heatmap’ values between 0 and 1; see left scale).

Since Figure 1 shows clearly contrasting blocks for the three patient groups along the diagonal based on the extracted relevant genes, these genes can be used to reliably classify the large majority of CG, RA, OA patients. It can be seen that only few patients, e.g. the last patient in the RA group (no. 23) and the last one in the OA group (no. 33), can not be reliably classified based on these genes.

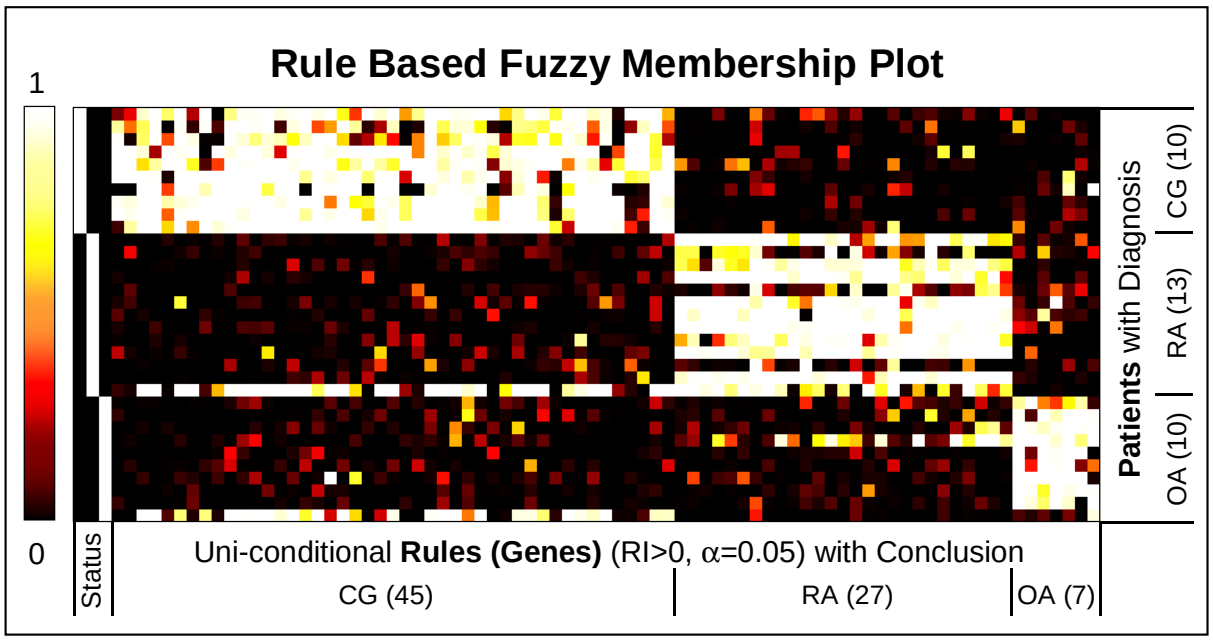


Figure 1: Rule based fuzzy membership plot for 33 patients and all 79 relevant genes

More specifically, the results are as follows:

Classification using the rulebase with all 79 extracted relevant rules (45 for CG, 27 for RA, 7 for OA) as shown in Figure 1 yielded a modelling error of 9%. 3 out of the 33 patients were incorrectly classified: no. 23 (actually RA, classified CG), no. 27 (actually OA, classified RA) and no. 33 (actually OA, classified CG), i.e. specificity was 100% and overall sensitivity 91%, with RA sensitivity 92% and OA sensitivity 80%.

Classification using an optimised rulebase with the 7 top ranked rules each for CG, RA, OA, i.e. altogether only 21 rules (Figure 1: the 7 rules furthest left for CG and for RA and all 7 rules for OA) resulted in a modelling error of just 3%. Now only 1 patient out of the 33, no. 33 (actually OA), was incorrectly classified as CG. Specificity here was also 100%, while overall sensitivity improved to 96%, with RA sensitivity improved to 100% and OA sensitivity to 90%.

The results obtained using the optimised rulebase were validated by leave-one-out cross-validation. This yielded a generalisation error of 24%. The incorrect classifications (8) occurred when patients no. 7 (actually CG), no. 12, 15, 21, 23 (all actually RA) and no. 27, 30, 33 (all actually OA) were left out. Among these 8 patients were those 3 who were, in the same way, also incorrectly classified when the original rulebase was used (no. 23, 27, 33). The additional 5 incorrectly classified patients were no. 7 (actually CG, classified OA), no. 12, 15, 21 (all actually RA, all classified OA) and no. 30 (actually OA, classified CG).

In summary, the study established a viable classification approach for this particular kind of gene expression data in conjunction with clinical diagnosis data employing a blend of unsupervised methods (data pre-processing and clustering) and supervised methods (rule extraction and rulebase construction). Although the analysis of 33 patients already provided good classification results, further validation with additional gene expression data for larger numbers of patients is required to finally establish gene expression based decision support systems for the differential diagnosis of rheumatic diseases.

Authors:

Dr Michael Pfaff [1], Dirk Woetzel [1], Dominik Driesch [1], Dr Susanne Toepfer [1], Dr René Huber [2], Dr Dirk Pohlers [2], Dr Dirk Koczan [3], Prof Hans-Juergen Thiesen [3], Dr Reinhard Guthke [4], Prof Raimund W. Kinne [2]

[1] BioControl Jena GmbH
Wildenbruchstr. 15
D-07745 Jena, Germany
Phone: +49-3641-527831
Fax: +49-3641-527832
E-mail: michael.pfaff@biocontrol-jena.com

[2] Friedrich Schiller University Jena
Medical Faculty
Department of Orthopaedics
Experimental Rheumatology Unit
Waldkrankenhaus 'Rudolf Elle' gGmbH
Klosterlausnitzer Str. 81
D-07607 Eisenberg, Germany

[3] University of Rostock
Medical Faculty
Institute for Immunology
Schillingallee 70
D-18055 Rostock, Germany

[4] Leibniz Institute for Natural Product Research
and Infection Biology e.V. - Hans Knoell Institute
Department of Molecular and Applied Microbiology
Beutenbergstr. 11a
D-07745 Jena, Germany